High-throughput automated refolding screening of inclusion bodies

RENAUD VINCENTELLI,[1] STÉPHANE CANAAN,[1] VALÉRIE CAMPANACCI,
CHRISTEL VALENCIA,[2] DAMIEN MAURIN, FRÉDÉRIC FRASSINETTI,
LORÉNA SCAPPUCINI-CALVO, YVES BOURNE, CHRISTIAN CAMBILLAU,
AND CHRISTOPHE BIGNON

Architecture et Fonction des Macromolécules Biologiques, Unité Mixte de Recherche (UMR) 6098, Centre National
de la Recherche Scientifique (CNRS) et Universités d'Aix-Marseille I et II, 13402 Marseille Cedex 20, France

Abstract

One of the main stumbling blocks encountered when attempting to express foreign proteins in *Escherichia coli* is the occurrence of amorphous aggregates of misfolded proteins, called inclusion bodies (IB). Developing efficient protein native structure recovery procedures based on IB refolding is therefore an important challenge. Unfortunately, there is no "universal" refolding buffer: Experience shows that refolding buffer composition varies from one protein to another. In addition, the methods developed so far for finding a suitable refolding buffer suffer from a number of weaknesses. These include the small number of refolding formulations, which often leads to negative results, solubility assays incompatible with high-throughput, and experiment formatting not suitable for automation. To overcome these problems, it was proposed in the present study to address some of these limitations. This resulted in the first completely automated IB refolding screening procedure to be developed using a 96-well format. The 96 refolding buffers were obtained using a fractional factorial approach. The screening procedure is potentially applicable to any nonmembrane protein, and was validated with 24 proteins in the framework of two Structural Genomics projects. The tests used for this purpose included the use of quality control methods such as circular dichroism, dynamic light scattering, and crystallogenesis. Out of the 24 proteins, 17 remained soluble in at least one of the 96 refolding buffers, 15 passed large-scale purification tests, and five gave crystals.

Keywords: screening; refolding; solubility; inclusion bodies; automation; high-throughput

In the context of Structural Genomics (SG) projects involving targets from *Escherichia coli* (ASG), *Mycobacterium tuberculosis* (MT), and viruses (SPINE), we have performed expression assays on ~600 genes (Sulzenbacher et al. 2002; Vincentelli et al. 2003). One of the main obstacles we and other authors have encountered when expressing recombinant proteins in *E. coli* is the relatively low soluble protein yield obtained with many of the source organisms used. In the case of eukaryotes, viruses, and *Mycobacterium tuberculosis*, most of the genes were expressed in the form of insoluble aggregates called "inclusion bodies" (IB). This obstacle to obtaining suitable targets for performing structural studies was particularly severe in the case of MT, with which 93% of our 182 tar-

gets yielded IB when proteins were expressed fused to an N-terminal His tag.

IBs are assumed to result from illegitimate interactions between hydrophobic residues located in the core of different molecules. This process is auto-catalyzed and therefore rapidly results in the precipitation of all the recombinant proteins produced in the cell (Mukhopadhyay 1997). Methods have been designed to recover correctly folded proteins from these amorphous aggregates. These include the "dilution," "dialysis," and "solid phase" methods (De Bernardez-Clark 1998), all of which involve an initial IB solubilization step using highly concentrated solutions of chaotropic agents such as guanidinium chloride and urea. The subsequent step in all these methods consists of removing the denaturing agent and restoring the protein to its native shape from the unfolded soluble state. The pathway used to remove the chaotropic agent differs between the three methods, however, although the same result is reached in each case. With the dilution method, refolding is assumed to occur immediately upon diluting the protein in a large volume of nondenaturing buffer ("refolding buffer"), which has to be sufficiently large to both cancel out the solubilizing effect of the chaotropic agent and reduce the probability that protein interactions will occur. The dialysis method involves the use of the same initial and final buffer compositions as the dilution method, but in this case, there is no dilution to decrease the protein–protein contacts (Rudolph and Lilie 1996; Mukhopadhyay 1997). Finally, it was established that physically separating molecules from each other during the renaturation process (solid phase refolding) greatly improved the refolding yield (Stempfer et al. 1996).

Whatever the method used to replace denaturing by nondenaturing buffer (a dilution, dialysis, or solid state method), it would be easier to use a single refolding buffer. Unfortunately, experience has shown that the composition of the refolding buffer is strongly protein dependent and that simply maintaining a difference between the pH of the refolding buffer and the isoelectric point (IP) of the protein does not usually suffice to keep the protein soluble.

Hence the idea of testing several refolding buffers simultaneously. For instance, Perbio has addressed this issue with Pro-Matrix, a refolding kit consisting of nine basic buffers, which can be supplemented with additives (Qoronfleh 2004). Using a fractional factorial approach, Armstrong et al. (1999), Chen and Gouaux (1997), and Hampton Research (FoldIt) have each developed separate procedures using 16 refolding conditions.

Despite these improvements, some difficulties were still encountered in the protein solubility assays performed to monitor the refolding process. Because no solubility assay was provided with the Pro-Matrix kit, this assay had to be set up by the customer, and the methods suggested for a solubility assay in the case of the FoldIt kit (size exclusion chromatography [SEC]), as well as those used by Armstrong et al. (1999) and Chen and Gouaux (1997) (dialysis and centrifugation), were not compatible with a high-throughput or with automation, which are two of the most crucial features in SG studies.

To solve the problems associated with the above limitations, a protein solubility test based on light scattering has been devised (Trésaugues et al. 2004). In practice, the turbidity of the solution is assessed by measuring the optical density (OD) at 390 nm, before and after adding the protein. If the protein remains soluble, the absorbance remains unchanged. In the opposite case, the OD increases proportionally to the amount of precipitate produced. This procedure is much faster than SEC and can be easily automated, but the number of conditions was still limited to 12, and the proteins often precipitated in all of them. This clearly suggested that the number of conditions needed to be further increased. A method of making this quantitative jump has been experimented in microtiter plate format, using 203 refolding conditions (Sijwali et al. 2001). However, the latter study was only designed for screening different GSH:GSSG ratios.

It is worth noting that although increasing the number of refolding conditions increases the probability that a protein will meet a buffer composition favoring its solubility, it also increases the number of samples to be handled. One possible solution to this problem consists of automating the screening process. In addition, automation is required to obtain sufficiently large SG throughputs. A partially automated refolding screening procedure was recently described (Scheich et al. 2004). With this procedure, however, the automation did not include any test for assessing the solubility and only 30 refolding conditions were used.

We therefore designed a refolding strategy involving the use of 96 different buffers in microtiter plate format, based on the above mentioned idea that the probability of a protein encountering a buffer composition favoring correct folding was likely to increase with the number of buffers tested. The solubility assay used in our screening procedure is basically the same as that described by Trésaugues et al. (2004), which accounts for protein *solubility*, and not for protein *folding*. After the preparatory refolding stage, circular dichroism (CD), dynamic light scattering (DLS), and crystallogenesis quality control procedures were added to respectively assess the folding, aggregation state, and homogeneity of the protein solution. These methods were chosen because they can be applied in theory to any protein, which is a prerequisite in the field of post-Genomics, which deals mainly with proteins having an unknown function. Finally, the availability of a pipetting robot made it possible to automate the whole process in a 96-well plate format.

To the best of our knowledge, this is the first completely automated "wide spectrum" 96-well IB refolding screening procedure to be developed based on a factorial approach. The present article describes the setup involved and con-

firms the validity of the method, based on tests carried out with proteins originating from two SG projects.

Results

Optimization of the solubility assay

The recently described solubility test, in which the turbidity of the solution is measured in terms of the light absorbance at 390 nm, involves light scattering by a protein precipitate (Trésaugues et al. 2004). As no proof was available that this wavelength was the most suitable one, we first addressed this point.

For this purpose, the absorbance of a bovine serum albumin (BSA) precipitate was scanned between 230 and 600 nm. As shown in Figure 1 (curve A), the absorbance decreased continuously from 230 to 600 nm. In addition to this regular decay, a small shoulder was present in the 280 nm region. To determine whether this feature was due to any remaining soluble proteins, the precipitate was spun down and the scanning performed again on the supernatant. Surprisingly, in this case, $OD_{230-600}$ was indistinguishable from the baseline, which means that the protein content had been entirely converted into insoluble species. These results indicate that the absorbance pattern of the protein precipitate, which is shown in Figure 1 (curve A), was entirely accounted for in terms of light scattering and not even partially in terms of the absorbance of soluble proteins.
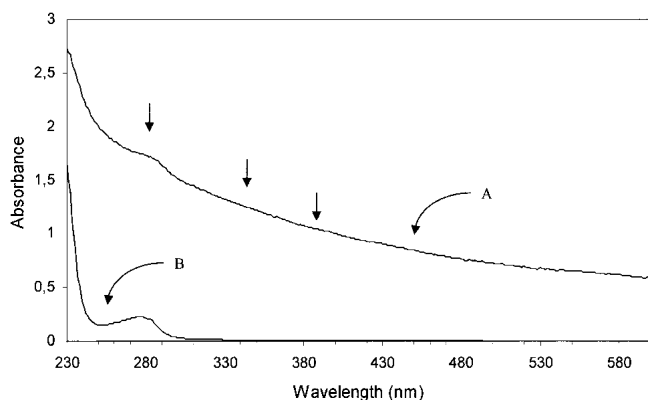
Figure 1. Absorbance spectra of precipitated and soluble forms of a protein. Twenty microliters of a 20 mg/mL BSA solution were diluted in 500 μL of either 100% isopropanol or 8 M guanidinium chloride. A chaotropic solution was used to ensure that the entire protein content was soluble. The absorbance of the resulting protein suspension (in isopropanol) or solution (in guanidinium chloride) was recorded from 230 to 600 nm, using a Varian Cary Scan 50 spectrophotometer. After subtracting the baseline (the absorbance of each solvent in the absence of protein), the absorbance intensities were plotted vs. the wavelengths. (Curve *A*) Precipitated protein in isopropanol. (Curve *B*) Soluble protein in guanidinium chloride. From *left* to *right*, three vertical arrows indicate the position of 280, 340/350, and 390 nm wavelengths, respectively.

Because the solubility assay was expected to distinguish between the absorbance due to precipitated and soluble proteins, the same experiment was performed under conditions where the proteins remained 100% soluble. In this case (Fig. 1, curve B), the absorbance profile was that of a typical protein solution, peaking at 280 nm (aromatic side chains) and at 200 nm (peptide bonds). Note that only the beginning of the peptide bonds' absorbance peak (λ max 200 nm; Stoscheck 1990) was visible between 230 and 240 nm.

In conclusion, the wavelength to be used in the solubility test should satisfy the following contradictory criteria: (1) It should be high enough above 280 nm to prevent any risk of obtaining false negative results due to the absorbance of (partially or totally) soluble proteins, at values of 280 nm and below, but (2) it should be as small as possible to provide the highest signal-to-noise ratio, according to curve A, and hence the most sensitive assay. In practice, 340-(manual procedure) and 350-nm (automated procedure) wavelengths were selected because they fulfilled these two criteria and provided better results than 390 nm.

Selection of 96 refolding conditions

The chemicals listed in Table 1, which were used to prepare the refolding mixes presented in Figure 2, were selected on the basis of the following criteria:

1. A 4 pH to 9 pH range was chosen because the proteins to be screened had various IPs and were likely to denature below or above these values.

2. Various ionic strengths (none; 100 mM NaCl or KCl; and 200 mM NaCl) were used because the solubility can increase (salting in) or decrease (salting out) with the salt concentration from one protein to another.

3. With the dilution method used, refolding was allowed to proceed for a very short time. Amphiphilic components (glycerol, PEG) were introduced to prevent the hydrophobic residues of different molecules still accessible at intermediate refolding stages from interacting with each other. In addition, glycerol and PEG were already provided in other refolding kits (Trésaugues et al. 2004) and were compatible with crystallogenesis. Glucose and arginine were used for the same reason, although Arg had to be removed before the crystallogenesis trials (see below).

4. Solubilizing reagents in the NDSB series were selected because they have been successfully used in protein crystallogenesis (Karaveg et al. 2003) and refolding experiments (Vuillard et al. 1998; Expert-Bezancon et al. 2003).

5. Proteins bearing odd numbers of cystein can form unnatural intermolecular disulfide bonds, which is a pos-

Table 1. *Chemicals used to make the 80 first refolding buffers*

Buffer (50 mM)	Ionic strength	Amphiphilic	Detergent (100 mM)	Reducing agent (10 mM)	Additive
NaAc, pH 4	NaCl 100 mM	Glycerol 20% (v/v)	NDSB 195	β-MSH	Arginine 800 mM
MES, pH 5	NaCl 200 mM	PEG 4000 0.05% (w/v)	NDSB 201		Glucose 500 mM
MES, pH 6	KCl 100 mM	PEG 400 0.05% (w/v)	NDSB 256		Cocktail[a]
TRIS, pH 7					EDTA 1 mM
TRIS, pH 8					
CHES, pH 9					

The concentrations indicated are those used before adding the protein.
[a] Consisted of 50 μM of each of the following: NADH, thiamine HCl, biotin, $CaCl_2$, $MgCl_2$, $CuSO_4$, $ZnCl_2$, $CoSO_4$, ADP, and $NiCl_2$.

sible cause of precipitation during the refolding process. Ten millimolar of β-MSH were introduced to prevent this mispairing.

6. The "cocktail" contained potential cofactors that might be required during the refolding process in the case of some proteins, whereas some other proteins tend to precipitate in the presence of divalent cations, hence the presence of EDTA.

7. The chaotrops (urea and guanidinium chloride) present in the commercial kits were discarded because they were liable to damage the robot's pipetting valves.

It was necessary to use a fractional factorial approach on the first 80 wells, because the combination of 20 chemicals would have resulted in too many experimental points (the full factorial design would have been 2560 combinations).

In the 16 remaining microplate wells, mini chaperones (a soluble form of GroEL; Altamirano et al. 1997) and redox components (GSSH, GSSG, DsbA) were combined, because the disulfide bond formation/reduction during the folding process itself has been found to be crucial (Wei et al. 1999). Details of each of the refolding conditions are given in Figure 2.

Figure 2. Detailed composition of each well in the refolding plate. (*) Tris (pH 8), NaCl 150 mM, EDTA. For details, see Table 1.

Testing of 96 refolding conditions

The 96-well screening procedure was tested on a panel of 24 proteins from two SG projects: MT (18 targets) and SPINE (6 targets). The results obtained are given in Table 2. Eleven out of the 18 MT targets (61%) and all the SPINE targets subjected to screening remained soluble under at least one of the 96 refolding conditions. In addition, except for MT target Rv1373 (buffer 57), all the responsive targets remained soluble in many buffers, which made it possible to choose the most suitable one(s) for the downstream steps such as crystallogenesis. In addition, the pH was not found to be a decisive parameter, because most of the targets remained soluble in a wide pH range, except Rv1525, Rv1515c, Rv0323c, and Rv2045, which remained soluble only at pH 4. Generally speaking, no particular buffer composition (pH, ionic strength, etc.) peaked more than the others, which suggests that the solution was always protein specific. The solubility yield at the production stage also appeared to be very high: 10 out of the 11 responsive MT targets (91%), and five out of the six responsive SPINE targets (83%) succeeded in passing the large-scale refolding and the first concentration steps. Only one SPINE (63) and two MT (Rv0323c and Rv1515c) targets were lost during the second concentration step following the gel filtration. In these particular cases, CD was nonetheless performed, but on protein solutions with concentrations too low for crystallogenesis.

Validity of the refolding screening procedure

Protein *solubility* and *folding* superimpose satisfactorily, but the overlap is not always 100%. We therefore tried to assess

Table 2. *(A) MT and SPINE targets remaining soluble in at least one refolding buffer and (B) summary of positive targets at each step*

A	Target	MW	Organism	Soluble in buffer [a]	Purification	IP	pH	CD	DLS	Crystal
	Rv2391	66	MT	39, 54, 57	57 (−Arg)	6.31	8	nd	nd	No
	Rv2392	30	MT	39, 49, 55, 56, 59, 61, 63, 64, 66	59	5.87	8	Ok	nd	Yes
	Rv1399c	36	MT	41, 44, 48, 49, 56, 59, 65, 66	41	4.38	7	Ok	M	Yes
	Rv1208	37	MT	41, 43, 48, 54, 56, 59, 63, 65, 66, 68, 69, 70, 74, 80	74	4.75	9	Ok	A	Yes
	Rv1373	40	MT	57	57 (−Arg)	6.36	8	nd	A	No
	Rv1564c	84	MT	41, 43, 44, 49, 56, 57, 59, 63, 66	41	4.95	7	Ok	D	No
	Rv1523	40	MT	4, 7, 10, 11, 12	4 (−glyc)	8.06	4	Ok	nd	No
	Rv1515c	36	MT	4, 5, 7, 10, 11, 12	4 (−glyc)	6.79	4	Ok	nd	No[b]
	Rv0323c	27	MT	2, 3, 4, 5, 9, 10, 11, 12	4 (−glyc)	5.81	4	Ok	nd	No[b]
	Rv2045c	59	MT	3, 4, 5, 6, 7, 10, 11, 12	4	7.67	4	Ok	nd	No
	Rv3487c	29	MT	2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22, 23, 24, 29, 45, 47, 49, 54, 57, 75	nd	8.85		nd	nd	nd
	SPINE 5	23	Sendai	10, 58, 59, 67, 73, 76	69	5.06	9	Ok	T	Yes
	SPINE 10	23	Measles	1, 2, 3, 4, 5, 6, 9, 10, 11, 12, 16, 17, 22, 24, 26, 32, 49, 54, 75, 78, 79	6	8.99	4	Ok	A	No
	SPINE 21	52	SFV	1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 21, 29, 31, 45, 49, 75, 78	4, 6	8.80	4	nd	A	No
	SPINE 22	53	SFV	2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 31, 45, 49, 57, 78, 79	nd	9.03		nd	nd	nd
	SPINE 23[c]	23	Human	All except 5, 6, 16, 17, 26, 42, 53, 61, 65, 76	33	8.68	6	Ok	D	Yes
	SPINE 63	23	HIV	1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 27, 28, 29, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 47, 48, 49, 51, 52, 54, 55, 57, 58, 66, 78	19	9.9	5	Ok	H	nd[b]

B	Target number	Responsive targets	Large-scale purification	CD OK	DLS OK	Crystal
MT	18	11	10	8	2	3
SPINE	6	6	5	4	3	2

(Target) The Rv nomenclature used was that of the MT genome (Cole et al., 1998; Camus et al. 2002). (MW) theoretical molecular weight (kDa). (IP) isoelectric point (taking into account the His tag when present). (pH) pH of the mix used for large-scale purification. (CD) ok, the protein fulfilled the criteria defined in Materials and Methods. (DLS) Only the main (>95%) population (M, D, etc.) was included in the table. (M) monomeric; (D) dimeric; (T) tetrameric; (H) Hexameric; (A) Aggregates (see Materials and Methods for details).
[a] The numbers refer to the buffers listed in Fig. 2 (1 = 1A, 2 = 1B … 9 = 2A, etc.). (−Arg), (−glyc) protein purification was performed using the buffer indicated devoid of arginine or glycerol, respectively.
[b] Lost during gel filtration or after the last concentration step.
[c] This target was not refolded from IB, but from a Ni eluate that precipitated just after elution.
(Target number) Number of targets subjected to refolding screening. (Responsive targets) Number of targets subjected to refolding screening that remained soluble in at least one refolding buffer. (DLS OK) DLS was taken to be satisfactory when the criteria defined in Materials and Methods were fulfilled.

the overlap in the case of proteins that were quantitatively refolded. In post-Genomics, one is often dealing with genes encoding proteins with an unknown function, and functional tests for each of the targets are frequently lacking. Therefore, depending on the targets, generic and/or specific methods can be used to assess the folding.

Generic methods

Circular dichroism (protein folding), dynamic light scattering (protein aggregation), and crystallogenesis (protein folding and dispersion homogeneity) were used for this purpose. Note that out of the 17 targets that reached the large-scale refolding stage, five could not be subjected to CD analysis either because of the presence of NDSB in the refolding buffer or because the amount of protein available was not sufficient. Crystallogenesis was also taken to be a valid folding criterion, because only properly folded proteins with an even aggregation state yield well-ordered crystals.

The results obtained with these three methods, which are summarized in Table 2, indicated upon CD analysis that all the targets that produced crystals also displayed folding features. This was so in the case of both MT (Rv2392, Rv1399c, Rv1208) and SPINE (targets 5 and 23). However, the opposite was not true: CD-positive MT targets Rv1564c Rv1523, Rv1515c, Rv0323c, and Rv2045 and SPINE target 10 did not produce crystals. Therefore, although the sole presence of secondary structures (β-sheet and/or α-helix) did not necessarily lead to successful crystallogenesis, its absence could be said to suggest a poor prognosis in terms of crystallogenesis, at least with this particular protein sample. By contrast, protein aggregation detected by DLS analysis seems to have a lower predictive value, because MT target Rv1208 produced crystals despite its aggregated state. Finally, the crystallization yield obtained with this procedure (five targets [36%]) was outstandingly high.

Specific method

Although the presence of secondary structures (CD), the lack of aggregates (DLS), and crystal growth argue in favor of correct folding, it is necessary to carry out more specific tests whenever possible. This was the case with Rv1399c. Because this target had been annotated as a putative lipase, a specific enzymatic assay was set up (Canaan et al. 2004). As illustrated in Table 3, the enzymatic activity could be measured after the refolding step, which provides evidence that our refolding screening procedure yields functional proteins, and not only soluble proteins. Two additional points are worth noting in Table 3: First, the refolding yield could be assessed, and turned out to be particularly high (50%). Second, 24 h after the refolding process, the total enzymatic activity was six times higher, which reflects the occurrence of a slow refolding process.

Scale up: Criteria for the choice of refolding buffer

Isoelectric point

As can be seen from Table 2, whenever possible, we chose conditions giving the largest difference in pH with the isoelectric point (IP) of the protein. Although we do not know how many proteins would remain soluble if a mixture with a pH near the IP was used, our choice actually resulted in 100% of the targets being successfully purified.

Compatibility with downstream steps

High concentrations of arginine sometimes artificially maintained proteins in the soluble state. Consequently, the removal of arginine often resulted in protein precipitation (not illustrated). In addition, due to its "anti-aggregation" effects (Umetsu et al. 2003), 800 mM Arg would have hampered crystallogenesis. We therefore tested the solubility of Rv2391 and Rv1373 in buffers with decreasing concentrations of Arg. Because these proteins remained soluble without any Arg, we decided to purify them in Arg-free buffer 57.

Pipetting a solution containing both a high protein concentration and 20% glycerol would lead to poor perfor-

Table 3. *Rv1399c refolding in the preparatory stage*

Step	Protein (mg)	Total activity (U)	Active protein (mg)	Specific activity (U/mg)	Refolding yield (%)
Ni⁺ affinity column and concentration	160	0	0	0	0
Dilution in refolding buffer	160	12,880	9.5	80.5	5.9
Dilution in refolding buffer (24 h later)	160	77,760	57.6	486	36
Freezing/thawing (before centrifugation)	92	96,600	71.5	1050	44.7
Freezing/thawing (after centrifugation)	80	108,000	80	1350	50

The enzymatic activity was measured as described (Canaan et al. 2004), one unit (U) of activity being defined as the hydrolysis of one micromole of substrate per minute. The amount of active protein was calculated by dividing the total activity recorded at each step by the maximum specific activity (1350 U/mg). The refolding yield was calculated by dividing the amount of active protein obtained in each step by the amount of starting material (160 mg eluted from the Ni affinity column).

mances of the Cartesian crystallization robot in the ~100 nL range. The same dilution technique was therefore used with glycerol as that described above in the case of Arg, with similar results and effects on large-scale purification.

It can therefore be said that although Arg and glycerol were helpful *during* the refolding step, they were no longer required *subsequently* to maintain the solubility of the protein, at least with these particular targets.

### *Choosing between manual and automated procedures*

If a small number of proteins have to be screened, the manual procedure is preferable, whereas a large number of targets (tens to hundreds) requires the use of an automated procedure. In this case, screening one plate takes only 5 min, and in its present form, the robot can process 27 plates

in 2 h 30 min without any human interventions. Thanks to the color code, the automated procedure, in addition to saving time, made it possible to display the results in a form that was easier to analyze than the manual procedure (Fig. 3B).

Discussion

IB refolding versus soluble expression in SG

To manage our SG programs, we have developed a general strategy based on several "screening rounds" of increasing complexity (Vincentelli et al. 2003). In the first round, targets are expressed using a single vector encoding an N-terminal His-tag fusion and a single *E. coli* strain. In the second round, eight *E. coli* strains are transformed by the
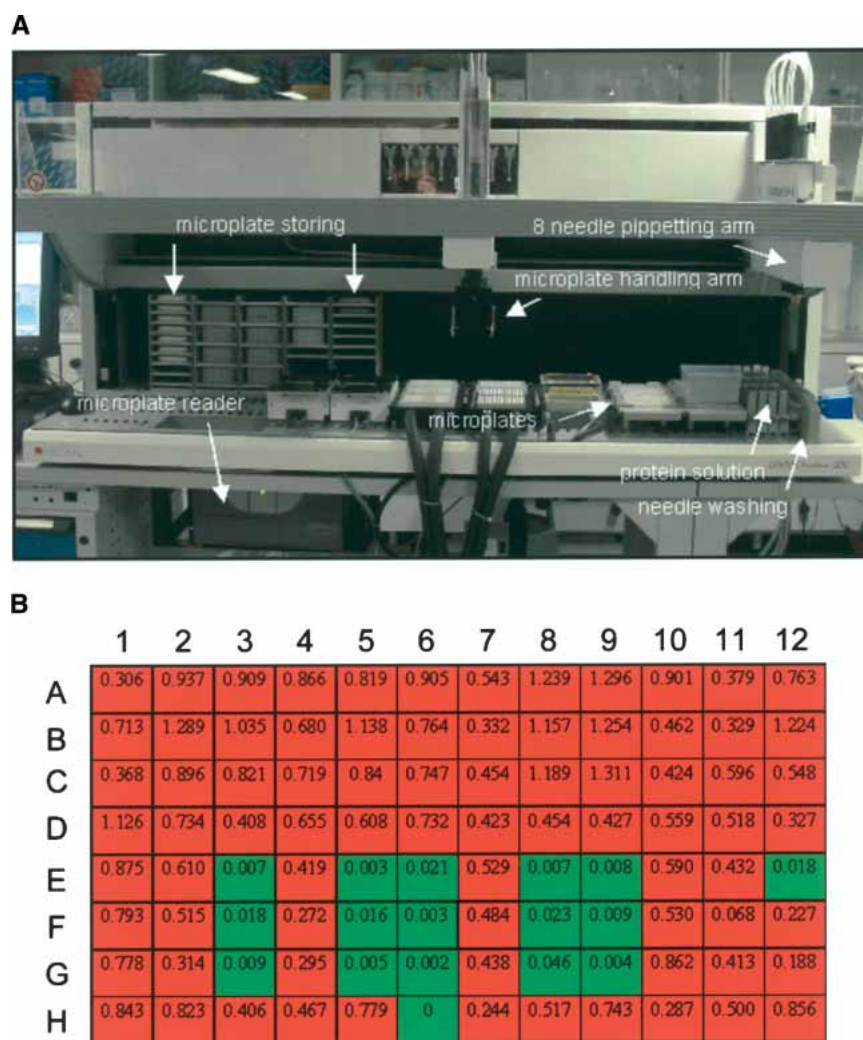
Figure 3. (*A*) Robot used in the automated procedure. The tools required for the refolding screening procedure are indicated by arrows. (*B*) Results of Rv2392 refolding screening. At the end of the experiment, the results (in Excel format) were displayed using a color code: Green and red indicate the wells containing soluble (DO < 0.05) and precipitated (DO > 0.05) proteins, respectively.

same vector as in round 1, and used to express the recombinant proteins at different temperatures. In the third round, the coding sequences are fused with maltose-binding protein, thioredoxin, glutathione S-transferase, and NusA. In the fourth round, the same experimental conditions are used as in round 1, except that the proteins are refolded from IB.

Comparisons between rounds 3 and 4

In the MT program, screening round 3 seems to be the most fruitful procedure so far, as it yielded 56 soluble proteins after proteolytic cleavage of the fusion (S. Canaan, R. Vincentelli, D. Maurin, F. Frassinetti, L. Scappucini-Calvo, Y. Bourne, C. Cambillau, and C. Bignon, unpubl.). However, its cost (in terms of the time required to prepare fusion constructs and to process the fusion vectors, the price of the endopeptidase, etc.) could easily be prohibitive. Conversely, IB refolding at preparative scale yielded 10 MT soluble proteins at a much lower cost, starting with only a fraction (27%) of the insoluble MT targets. Although no SPINE target was processed in round 3, it is worth noting that five out of six targets (83%) yielded soluble proteins in the preparatory stages of IB refolding, starting with only 3% of SPINE insoluble proteins.

Comparisons between rounds 1 and 4

In addition, the success rate (defined as the percentage of the proteins that succeeded in passing the scale-up step) obtained in round 4 with 18 MT and six SPINE targets (61% and 83%, respectively) was much more satisfactory than that obtained in round 1: Out of 182 MT and 244 SPINE target genes, only 14 (7.7%) MT targets and 80 (33%) SPINE targets were directly recovered in the form of soluble proteins after *E. coli* cell lysis. This means that at least in some cases, the IB chemical refolding procedure produces soluble species more efficiently than living bacteria. Therefore, we propose to adopt IB refolding in the initial stages of SG projects dealing with highly insoluble proteins, such as the MT project. The validity of this approach has been established in the case of small (<18 kDa) proteins intended for NMR structural analysis (Maxwell et al. 2003). Because 58% of the proteins were found to be properly refolded when a single renaturation buffer was used, one can expect to obtain a much higher refolding yield if an upstream refolding screening procedure is carried out in addition (Maxwell et al. 2003).

Limitations of the screening procedure

The 96-well plate refolding screening procedure is not suitable for use with either high pressure (St. John et al. 1999) or reverse micelle (Vinogradov et al. 2003) approaches, for physical reasons. Nor can this method be used to study refolding processes using time-dependent techniques such as stepwise dialysis with additives (Umetsu et al. 2003) or air oxidation techniques (Menzella et al. 2002). Other limitations of our method are due to the OD_{340} detection method used:

1. IB redissolved in chaotrop must be free of contaminants, otherwise these might promote precipitation, yielding false negative results. In this respect, the nickel affinity purification step is of particular importance.

2. If the protein concentration is too low in the chaotropic agent, there may be no detectable precipitate after diluting the protein in refolding buffer, even if the buffer is not favorable to maintaining the solubility.

3. We have observed that the first $OD_{340/350}$ reading was sometimes misleading: Some positive spots became negative due to the slow protein precipitation with time. The opposite also occurred, presumably due to the presence of proteins with slow refolding kinetics, such as Rv1399c (see Table 3). This prompted us to systematically perform a second $OD_{340/350}$ reading after a 1-d interval. Because refolded proteins must remain soluble throughout the long crystallization process, long-lasting solubility is more desirable than instant but transient solubility.

Possible improvements and perspectives

The high success rate obtained with the crystallogenesis procedure could be further increased by separating folded from misfolded species quantitatively prior to the crystallogenesis trials. Ion exchange chromatography might be a suitable method for this purpose, because folded and unfolded soluble forms of the same protein do not have the same overall charge. Reversed phase and hydrophobic interaction chromatography (Scheich et al. 2004) might also be suitable methods, as misfolded proteins are assumed to display a larger number of hydrophobic residues on their surface. Another possibility is to add a freezing/thawing step. In the case of Rv1399c, this step was found to differentiate between folded and unfolded populations: The inactive (misfolded) enzymes precipitated after thawing (12 mg), whereas the active (correctly folded) enzymes remained soluble, resulting in a greater specific activity (Table 3).

Low refolding yields and poorly diffracting crystals could also be improved by appropriately tuning one of the basic conditions provided by our refolding kit, as done routinely in crystallogenesis.

Another possible way of improving our screening method is to extend the pH range below 4 and above 9, because a dramatic increase in protein solubility has been reported to occur when the pH increased from 11 to 12.5 during the refolding of human growth hormone (Patra et al. 2000).

To improve the throughput, the Tecan Genios+ microplate reader can be used with 384-well plates. This would increase the number of mixtures to be screened on the same surface fourfold. If an automated plate sealer, a robot-driven centrifuge, and carousels in a 4°C atmosphere were added in the immediate vicinity of the robot, the screening process could be run nonstop, and the process would be completely automated.

Refolding screening can be performed at any stage in protein production procedures involving protein solubility problems, as illustrated here in the case of SPINE target 23, which was not refolded from IB, but was resolubilized from a precipitate that formed after the protein has been eluted from the Ni affinity column (Table 2).

Lastly, the role of lysis buffers in protein precipitation processes has been investigated (Lindwall et al. 2000). Because many of the refolding buffers in Figure 2 proved to be highly effective for maintaining protein solubility, they could also be tested for use as lysis buffers along with a 96-well sonicator.

Materials and methods

Cell growth and lysis

All coding sequences were subcloned by recombination (Gateway, Invitrogen) into the pDEST17O/I expression vector, a modified pDEST17 (Invitrogen) to which LacO and LacI were inserted to allow a better control of protein expression (Canaan et al. 2004). BL21(DE3)pLysS cells (Novagen) were transformed with 150 ng of the resulting constructs, and plated on ampicillin (100 μg/mL) and chloramphenicol (34 μg/mL). After one night at 37°C, all the colonies were scraped off the plate using a toothpick, and used to inoculate 1 L of LB. When the cell culture had reached an OD_{600} of 0.5, IPTG was added to a final concentration of 2 mM. After 4 h of shaking at 37°C, cells were recovered by centrifugation, resuspended in 50 mL of lysis buffer (50 mM Tris at pH 8, 150 mM NaCl, 1 mM EDTA, 0.1% Triton X100, 1 mM PMSF, 0.25 mg/mL lysozyme) and frozen overnight at −80°C.

Processing of inclusion bodies

After thawing the cell suspension, DNAse and $MgSO_4$ were added at final concentrations of 10 μg/mL and 20 mM, respectively. The lysate was incubated for 30 min at 37°C (or until it was no longer viscous), and then spun for 30 min at 17,000g. After discarding the supernatant, the pellet was thoroughly resuspended in 50 mL of Tris buffer (50 mM Tris at pH 8, 150 mM NaCl), disrupted by sonicating it four times with a 15-sec pulse, and spun again. This washing procedure was repeated three times. The final pellet was solubilized in 20 mL of 50 mM Tris, 150 mM NaCl, 10 mM imidazole, 8 M guanidinium chloride. After a 30-min run at 17,000g, the supernatant was loaded onto a 5-ml Chelatin sepharose fast-flow column (Amersham Bioscience) preequilibrated with 50 mM Tris, 150 mM NaCl, 10 mM imidazole, and 8 M urea (buffer A). The column was washed with buffer A supplemented with 50 mM imidazole, and the recombinant protein was eluted with buffer A containing 250 mM imidazole. Protein purity and integrity were checked by SDS-PAGE. Elution fractions containing the protein of interest were pooled, and the imidazole was removed using a desalting column. Proteins were concentrated to at least 5 mg/mL and, if necessary, cysteines were reduced by incubating for 1 h in the presence of 10 mM β-mercaptoethanol.

Screening plates

The 96 refolding mixes (Table 1; Fig. 2) were handmade and packaged in 50-mL tubes. A 95 μL aliquot of each refolding solution was dispensed by the Tecan Genesis Freedom 200 robot (Fig. 3A) into each of the 96 wells of flat-bottom clear microplates (Greiner). Approximately 500 plates/50 mL tube could be prepared in advance using this procedure. After filling the plates, they were sealed manually and stored at −20°C until use (we never observed any buffer precipitation at thawing).

Buffer compositions of wells A1 to H10 (see Fig. 2) were determined using SAmBA software (Audic et al. 1997). Wells A11 to H12 were filled with a single buffer (50 mM Tris at pH 8, 150 mM NaCl, 1 mM EDTA) containing different combinations of chaperones and/or redox components (DsbA, dbGroEL, GSH, GSSG).

Refolding screening

Manual and robot-assisted procedures have been set up.

Manual procedure

After thawing, the content of each of the 96 wells was mixed individually with 5 μl of an ~5 mg/mL urea-denatured and β-MSH-reduced protein solution (see "Processing of inclusion bodies," above), using a multichannel pipette. Immediately after the mixing step, the turbidity was assessed by measuring the OD at 340 nm using a μQuant microplate reader (BioTek Instruments Inc.) and the KC4 software program. The blank, that is the absorbance before adding the protein, was automatically subtracted by the computer program by recalling this previously recorded data. The protein was taken to be soluble at OD < 0.05. The plate was sealed and then stored at 4°C. Twenty-four hours later, the seal was removed and a second reading was performed.

Automated procedure

A Tecan Genesis Freedom 200 robot with an eight-needle pipetting arm, a microplate handling arm, and a Tecan Genios Plus microplate reader was used (Fig. 3A). The robot transferred 5 μL of an ~5 mg/mL urea-denatured and β-MSH-reduced protein solution from a microtube into each of the 96 wells. After dispensing the solution, the robot moved the microplate into the reader. The latter mixed the contents of each well by shaking the plate for 30 sec, and then measured the optical density at a wavelength of 350 nm (and not 340 nm as in the manual procedure, because this filter was not available for the automated setup). At the end of the assay, the software driving the reader (Magellan) subtracted the previously recorded blank value, and the resulting OD value could be displayed on the computer screen in the form of a microplate layout with a color code. At OD < 0.05 the well was green, and at OD > 0.05 the well background was red (see Fig. 3B). Because the robot could not seal or store the plates at 4°C, this was done manually. On the next day, the plates were returned to the robot and, after stripping the cover, another reading at 350 nm was performed. As expected, the wavelength shift from 340 to 350 nm resulted in only negligible differences (see Fig. 1).

Large-scale protein refolding

The total volume of the denatured protein obtained at the end of the "inclusion body processing" step was diluted 20-fold in the refolding buffer selected at the end of the screening procedure. When more than one buffer could be used, the default choice was that exhibiting the highest compatibility with the downstream steps (CD, DLS, crystallogenesis). If the first buffer selected proved to be unsatisfactory at any of the subsequent steps, a second one was tested, and so on until a buffer that was usable throughout was found. The renatured protein was then concentrated in a stirred Amicon cell and further purified by SEC using the same refolding buffer. This step removed the remaining 0.4 M urea still present after diluting the protein in the refolding buffer, along with any unwanted compounds (see Results section). The protein was concentrated to 5 mg/mL and tested by performing parallel CD, DLS analysis, and automated crystallogenesis trials.

Circular dichroism, dynamic light scattering, crystallogenesis

When the refolding buffer composition was compatible (i.e., when the buffer did not absorb between 190 and 260 nm), the presence of secondary structures was assessed by performing CD analysis (Jasco PTC-423S). Data deconvolution by the CDNN program provided the percentages of the strands, helices, turns, and random coils, which varied with both the molecular mass and the concentration used. Spectra of purified protein (final concentration 0.2 mg/mL) were recorded at 20°C at wavelengths ranging between 190 and 260 nm, with a 30-min averaging step. The final CD spectrum obtained was the mean of three measurements. The protein was taken to have significant secondary structure features when the α-helix and β-sheets amounted to more than 30%.

In addition to promoting secondary structures, efficient refolding buffers were expected to favor mono- (or pauci-) meric states of the refolded protein. To assess the state of aggregation of the proteins after the refolding process, these were subjected to DLS analysis in line with the manufacturer's instructions. Experiments were performed with a Dynapro MSTC-200 (Protein Solutions) at 20°C. Samples were filtered prior to the measurements (using Millex syringe filters, pore size 0.22 μm; Millipore Corp.). The fractions of monomer, dimer, and so forth were calculated using the software program provided by the manufacturer. Proteins were taken to be aggregated when the hydrodynamic radius was ≥10 nm (in general, the R_h of proteins is in the 2–4-nm range). The result of a DLS experiment was taken to be acceptable when (1) the polydispersity was moderate (<30%), (2) the major component did not consist of aggregates, and (3) the major component comprised at least 95% of the detected particles.

Crystallogenesis was performed using nano-drop robotics, as previously described (Sulzenbacher et al. 2002).

References

Altamirano, M., Golbik, R., Zahn, R., Buckle, A., and Fersht, A. 1997. Refolding chromatography with immobilized mini-chaperones. *Proc. Natl. Acad. Sci.* **94:** 3576–3578.

Armstrong, N., de Lencastre, A., and Gouaux, E. 1999. A new protein folding screen: Application to the ligand binding domains of a glutamate and kainate receptor and to lysozyme and carbonic anhydrase. *Protein Sci.* **8:** 1475–1483.

Audic, S., Lopez, F., Claverie, J., Poirot, O., and Abergel, C. 1997. SAmBA: An interactive software for optimizing the design of biological macromolecules crystallization experiments. *Proteins* **29:** 252–257.

Camus, J., Pryor, M., Medigue, C., and Cole, S. 2002. Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* **148:** 2967–2973.

Canaan, S., Maurin, D., Chahinian, H., Pouilly, B., Durrousseau, C., Frassinetti, F., Scappucini-Calvo, L., Cambillau, C., and Bourne, Y. 2004. Expression and characterization of the protein Rv1399c from Mycobacterium tuberculosis: A novel carboxyl esterase structurally related to the HSL family.

Chen, G. and Gouaux, E. 1997. Overexpression of a glutamate receptor (GluR2) ligand binding domain in *Escherichia coli*: Application of a novel protein folding screen. *Proc. Natl. Acad. Sci.* **94:** 13431–13436.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry 3rd, C.E., et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393:** 537–544.

De Bernardez-Clark, E. 1998. Refolding of recombinant proteins. *Curr. Opin. Biotechnol.* **9:** 157–163.

Expert-Bezancon, N., Rabilloud, T., Vuillard, L., and Goldberg, M. 2003. Physical-chemical features of non-detergent sulfobetaines active as protein-folding helpers. *Biophys. Chem.* **100:** 469–479.

Karaveg, K., Liu, Z., Tempel, W., Doyle, R., Rose, J., and Wang, B. 2003. Crystallization and preliminary X-ray diffraction analysis of lectin-1 from *Pseudomonas aeruginosa*. *Acta Crystallogr. D Biol Crystallogr.* **59:** 1241–1242.

Lindwall, G., Chau, M., Gardner, S., and Kohlstaedt, L. 2000. A sparse matrix approach to the solubilization of overexpressed proteins. *Protein Eng.* **13:** 67–71.

Maxwell, K., Bona, D., Liu, C., Arrowsmith, C., and Edwards, A. 2003. Refolding out of guanidine hydrochloride is an effective approach for high-throughput structural studies of small proteins. *Protein Sci.* **12:** 2073–2080.

Menzella, H., Gramajo, H., and Ceccarelli, E. 2002. High recovery of prochymosin from inclusion bodies using controlled air oxidation. *Protein Expr. Purif.* **25:** 248–255.

Mukhopadhyay, A. 1997. Inclusion bodies and purification of proteins in biologically active forms. *Adv. Biochem. Eng. Biotechnol.* **56:** 61–109.

Patra, A., Mukhopadhyay, R., Mukhija, R., Krishnan, A., Garg, L., and Panda, A. 2000. Optimization of inclusion body solubilization and renaturation of recombinant human growth hormone from *Escherichia coli*. *Protein Expr. Purif.* **18:** 182–192.

Qoronfleh, W. 2004. "Misfolded proteins? Inclusion bodies? Restoring protein function with the Pro-Matrix protein refolding system." Paper presented at the PEPTALK-Protein Expression meeting, San Diego, CA. http://www.chi-peptalk.com.

Rudolph, R. and Lilie, H. 1996. In vitro folding of inclusion body proteins. *FASEB J.* **10:** 49–56.

Scheich, C., Niesen, F., Seckler, R., and Bussow, K. 2004. An automated in vitro protein folding screen applied to a human dynactin subunit. *Protein Sci.* **13:** 370–380.

Sijwali, P., Brinen, L., and Rosenthal, P. 2001. Systematic optimization of expression and refolding of the *Plasmodium falciparum* cysteine protease falcipain-2. *Protein Expr. Purif.* **22:** 128–134.

St. John, R., Carpenter, J., and Randolph, T. 1999. High pressure fosters protein refolding from aggregates at high concentrations. *Proc. Natl. Acad. Sci.* **96:** 13029–13033.

Stempfer, G., Holl-Neugebauer, B., and Rudolph, R. 1996. Improved refolding of an immobilized fusion protein. *Nat. Biotechnol.* **14:** 329–334.

Stoscheck, C.M. 1990. Guide to protein purification: Quantitation of protein. *Methods Enzymol.* **182:** 50–68.

Sulzenbacher, G., Gruez, A., Roig-Zamboni, V., Spinelli, S., Valencia, C.,

Pagot, F., Vincentelli, R., Bignon, C., Salomoni, A., Grisel, S., et al. 2002. A medium-throughput crystallization approach. *Acta Crystallogr. D Biol. Crystallogr.* **58:** 2109–2115.

Trésaugues, L., Collinet, B., Minard, P., Henkes, G., Aufrère, R., Blondeau, K., Liger, D., Zhou, C.-Z., Janin, J., van Tilbeurgh, H., et al. 2004. Refolding strategies from inclusion bodies in a structural genomics project. *J. Struct. Funct. Gernomics* **5:** 195–204.

Umetsu, M., Tsumoto, K., Hara, M., Ashish, K., Goda, S., Adschiri, T., and Kumagai, I. 2003. How additives influence the refolding of immunoglobulin-folded proteins in a stepwise dialysis system. Spectroscopic evidence for highly efficient refolding of a single-chain Fv fragment. *J. Biol. Chem.* **278:** 8979–8987.

Vincentelli, R., Bignon, C., Gruez, A., Canaan, S., Sulzenbacher, G., Tegoni, M., Campanacci, V., and Cambillau, C. 2003. Medium-scale structural genomics: Strategies for protein expression and crystallization. *Acc. Chem. Res.* **36:** 165–172.

Vinogradov, A., Kudryashova, E., Levashov, A., and van Dongen, W. 2003. Solubilization and refolding of inclusion body proteins in reverse micelles. *Anal. Biochem.* **320:** 234–238.

Vuillard, L., Rabilloud, T., and Goldberg, M. 1998. Interactions of non-detergent sulfobetaines with early folding intermediates facilitate in vitro protein renaturation. *Eur. J. Biochem.* **256:** 128–135.

Wei, C., Tang, B., Zhang, Y., and Yang, K. 1999. Oxidative refolding of recombinant prochymosin. *Biochem. J.* **340:** 345–351.